Supplement to Semantic integration of physiology phenotypes with an application to the Cellular Phenotype Ontology

Robert Hoehndorf[1]*, Midori A. Harris[2], Heinrich Herre[3], Gabriella Rustici[4] and Georgios V. Gkoutos[1]

[1]Department of Genetics, University of Cambridge, Downing Street, Cambridge, Cambridge CB2 3EH, UK
[2]Department of Biochemistry; University of Cambridge, 80 Tennis Court Road, Cambridge CB2 1GA, UK
[3]Institute for Medical Informatics, Statistics and Epidemiology, University of Leipzig, Haertelstrasse 16-18, 04107 Leipzig, Germany
[4]European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, Cambridge CB10 1SD, UK

1 Axiom patterns

1.1 Basic pattern

Our basic pattern for formalizing phenotypes is derived from previous work [3]. We use the relation **phenotype-of** as a relation between a phenotype and some entity (the bearer of the phenotype). We assume that the bearer of the phenotype is an organism. Using the **phenotype-of** relation, we further specify the bearer's properties axiomatically. For example, a phenotype *Broad forehead* is a phenotype of an entity (the organism) that has a nose as part which has a quality of being *Broad*:

```
'Broad forehead' EquivalentTo:
  phenotype-of some (has-part some
    Nose and has-quality some Broad)
```

*to whom correspondence should be addressed

1.2 Comparison to normal

Phenotypes are often comparative to some reference that is considered as *normal* within a context, while the phenotypes specify deviations from this normality. One possibility to express deviation from normality is to explicitly contextualize the classes that represent the reference entity and either explicitly state or infer that entities with abnormal phenotypes are distinct from the reference entities [3]. Being abnormal is then equivalent to "not being normal". Through the explicit use of the **has-part** relation, it becomes possible to formally include *absence* as a type of abnormality: an *Abnormality of X* would be equivalent to *not having a normal X as part*, while the *Absence of X* is equivalent to *not having X as part*. The resulting inference, that *Absence of X* is a subclass of *Abnormality of X*, is widely implemented in phenotype ontologies and should be accommodated by our axiom patterns.

However, the explicit contextualization of classes with *normal* and *abnormal* qualifiers requires the use of disjunction, and the formalization of *Abnormality of X* as *not having a normal X as part* requires the use of negation. Neither operation is supported in the OWL EL profile [6], and since we limit ourselves to the OWL EL profile due to its low computational complexity, we cannot implement these patterns. As an alternative, we chose to implement abnormality and absence following formalization patterns already implemented in phenotype ontologies [7, 4], since these patterns can be expressed in OWL EL. As results, we can perform automated reasoning over our formalizaton and benefit from interoperability with previous work on formalizing phenotype ontologies. On the other hand, the use of these patterns may lead to unintended consequences for some queries [1]. Furthermore, although we could distinguish between *normal* and *abnormal* properties of entities using the PATO framework, we can also omit this distinction when we build an ontology that is solely composed of abnormal phenotypes or in which entities only have abnormal attributes. Based on these considerations, an example formalization for *Abnormality of cell cycle* is

```
'Abnormality of X' EquivalentTo:
  phenotype-of some ( has-part some
    ( part-of some 'Cell cycle' and
      has-quality some Quality))
```

The use of the class `part-of some 'Cell cycle'` instead of `'Cell cycle'` allows us to reuse the **part-of** relation to infer that abnormalities of parts of the cell cycle become subclasses of abnormalities of the cell cycle.

1.3 Definition of complex properties and classes

We use several relations in order to construct complex classes. In particular, we use the **regulates** relation from the GO to define classes of regulation processes, and we use the **towards** to specify a required second argument in some qualities. For example, we define the class *Decreased frequency of DNA ligation* as:

```
'Decreased frequency of DNA ligation' EquivalentTo:
```

```
  phenotype-of some (has-part some (participates-in some
    ((has-quality some
        ('decreased frequency'
         and (towards some 'DNA ligation')))
     and (regulates some 'DNA ligation'))))
```

This definition states that a *Decreased frequency of DNA ligation* is a phenotype of organisms that have a part that participates in a *Regulation of DNA ligation* process which has a *Decreased frequency of DNA ligation* (`'decreased frequency' and (towards some 'DNA ligation')`) as quality. In this definition, we refer to the class *Regulation of DNA ligation* based on its definition `regulates some 'DNA ligation'`, and to the class *Decreased frequency of DNA ligation* using the class description `'decreased frequency' and towards some 'DNA ligation'`. Although the use of the **towards** relation in combination with a quality such as *Decreased frequency* has several problems [5], it will lead to interoperability with the large number of phenotype ontologies that follow a similar pattern [7].

1.4 Chemical substances

We use the ChEBI ontology [2] to refer to classes that represent chemical entities or substances. In ChEBI, no distinction is made explicit between chemical entities, in the sense of individual molecules, and substances composed of these entities. In some applications, it can be necessary to distinguish between both in ontology-based information systems or tasks that react differently to both types (e.g., that distinguish between molecular weights and melting points) []. However, we reuse GO axioms that utilize ChEBI, and the GO does not make a distinction between the molecular entity and the substance composed of it. Similarly, we do not require such a distinction to achieve the aims of our axiomatization. Therefore, we implicitly treat classes in ChEBI as either molecular entities or substances composed thereof. If this distinction is made explicit, we can define each class that we currently use from ChEBI as a disjunction of the class representing a molecular entity and the substance composed of the entity.

1.5 Detailed axiom patterns

Here, we present the precise formulation of axiom patterns. We use X and A as variables in these patterns that are replaced with specific classes from the GO or ChEBI ontologies.

Phenotype of X: If X is a material object, we assert:

```
'X phenotype' EquivalentTo:
  phenotype-of some (has-part some
    (part-of some X))
```

If X is a cellular biological process, we assert:

```
'X phenotype' EquivalentTo:
  phenotype-of some (has-part some
    (participates-in some (part-of some X)))
```

Abnormality of X: If X is a material object:

```
'X phenotype' EquivalentTo:
  phenotype-of some (has-part some
    (part-of some (X and has-quality some Quality)))
```

If X is a process, we create a class that does not have any formal constraints. In CPO, abnormalities of X processes include both abnormalities of X and abnormality of regulation of X. A formulation of this constraint requires the use of disjunction. Since we limit ourselves to the use of OWL EL and disjunction is not a valid logical operator in OWL EL, we do not add any axiom to CPO in this case.

Abnormal morphology and physiology, and absence: Abnormal X morphology:

```
'X phenotype' EquivalentTo:
  phenotype-of some (has-part some
    (X and has-quality some Morphology))
```

Abnormal X physiology:

```
'X phenotype' EquivalentTo:
  phenotype-of some (has-part some
    (part-of some (X and has-quality some Functionality)))
```

Formalizing absence of X requires the use of negation. Since negation is not allowed in OWL EL, we create a class for absence that we directly assert to be a subclass of *Abnormality of X*, without adding further axioms.

Single occurrence abnormalities: Abnormality of single occurrence of X:

```
'Abnormality of single occurrence of X' EquivalentTo:
phenotype-of some (has-part some (
participates-in some (part-of some X and
has-quality some Quality)))
```

Increased duration of X:

```
'Increased duration of X' EquivalentTo:
phenotype-of some (has-part some (
participates-in some (X and
has-quality some 'Increased duration')))
```

Decreased duration of X:

```
'Decreased duration of X' EquivalentTo:
phenotype-of some (has-part some (
participates-in some (X and
has-quality some 'Decreased duration')))
```

Abnormalities of X regulation: Abnormality of regulation of X:

```
'Abnormality of regulation of X' EquivalentTo:
phenotype-of some (has-part some (
participates-in some (regulates some X and
has-quality some (Quality and towards some X))))
```

Increased frequency of X:

```
'Increased frequency of X' EquivalentTo:
phenotype-of some (has-part some (
participates-in some (regulates some X and
has-quality some ('Increased frequency' and towards some X))))
```

Decreased frequency of X:

```
'Decreased frequency of X' EquivalentTo:
phenotype-of some (has-part some (
participates-in some (regulates some X and
has-quality some ('Decreased frequency' and towards some X))))
```

Late onset of X:

```
'Late onset of X' EquivalentTo:
phenotype-of some (has-part some (
participates-in some (regulates some X and
has-quality some (Delayed and towards some X))))
```

Early onset of X:

```
'Early onset of X' EquivalentTo:
phenotype-of some (has-part some (
participates-in some (regulates some X and
has-quality some (Premature and towards some X))))
```

Abnormality of inputs and outputs (rates): If, by virtue of GO-XP definitions, we are able to identify inputs and outputs of X, we define the following classes:

Increased mass of A as output in regulation of X:

```
'Decreased mass of A as output in regulation of X' EquivalentTo:
phenotype-of some (has-part some (
participates-in some (regulates some X and
has-output some (A and has-quality some 'Increased mass'))))
```

Decreased mass of A as output in regulation of X:

```
'Decreased mass of A as output in regulation of X' EquivalentTo:
phenotype-of some (has-part some (
participates-in some (regulates some X and
has-output some (A and has-quality some 'Decreased mass'))))
```

Increased mass of A as output in X:

```
'Increased mass of A as output in X' EquivalentTo:
phenotype-of some (has-part some (
participates-in some (X and
has-output some (A and has-quality some 'Increased mass'))))
```

Decreased mass of A as output in X:

```
'Decreased mass of A as output in X' EquivalentTo:
phenotype-of some (has-part some (
participates-in some (X and
has-output some (A and has-quality some 'Decreased mass'))))
```

We assert similar axioms for the inputs of X.

References

[1] Martin Boeker, Ilinca Tudose, Janna Hastings, Daniel Schober, and Stefan Schulz. Unintended consequences of existential quantifications in biomedical ontologies. *BMC Bioinformatics*, 12(1):456, 2011.

[2] K. Degtyarenko, P. Matos, M. Ennis, J. Hastings, M. Zbinden, A. McNaught, R. Alcantara, M. Darsow, M. Guedj, and M. Ashburner. ChEBI: a database and ontology for chemical entities of biological interest. *Nucleic Acids Research*, 2007.

[3] Robert Hoehndorf, Anika Oellrich, and Dietrich Rebholz-Schuhmann. Interoperability between phenotype and anatomy ontologies. *Bioinformatics*, 26(24):3112 – 3118, 10 2010.

[4] Robert Hoehndorf, Paul N. Schofield, and Georgios V. Gkoutos. Phenomenet: a whole-phenome approach to disease gene discovery. *Nucleic Acids Research*, 39(18):e119, 2011.

[5] Frank Loebe, Frank Stumpf, Robert Hoehndorf, and Heinrich Herre. Towards improving phenotype representation in owl. In *Proceedings of the 3rd Workshop for Ontologies in Biomedicine and Life sciences (OBML)*, October 2011.

[6] Boris Motik, Bernardo Cuenca Grau, Ian Horrocks, Zhe Wu, Achille Fokoue, and Carsten Lutz. Owl 2 web ontology language: Profiles. Recommendation, World Wide Web Consortium (W3C), 2009.

[7] Christopher Mungall, Georgios Gkoutos, Cynthia Smith, Melissa Haendel, Suzanna Lewis, and Michael Ashburner. Integrating phenotype ontologies across multiple species. *Genome Biology*, 11(1):R2+, 2010.